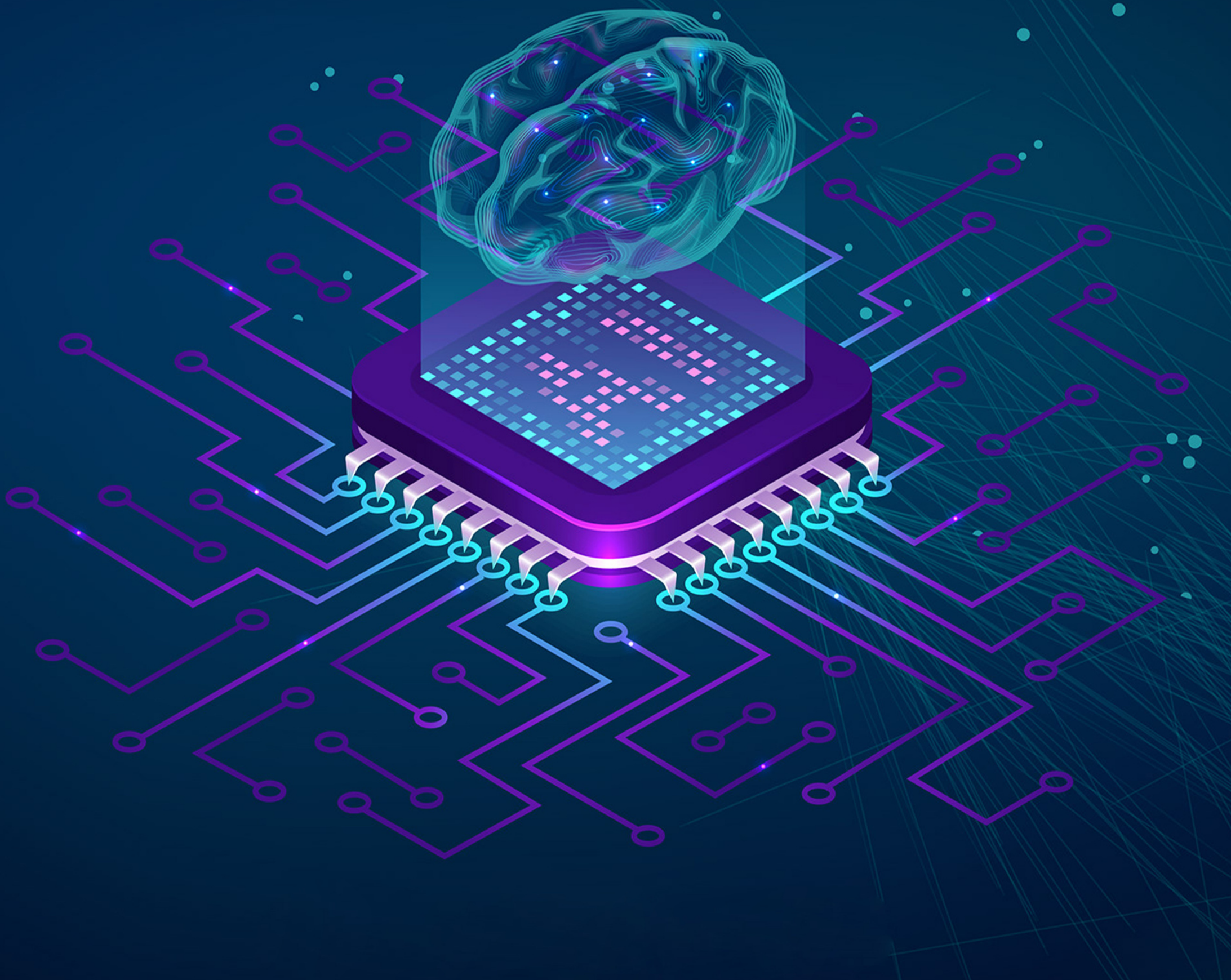
AlgoTutor

25 IMPORTANT DATA SCIENCE INTERVIEW QUESTIONS



AlgoTutor

Q 1. What is data science?

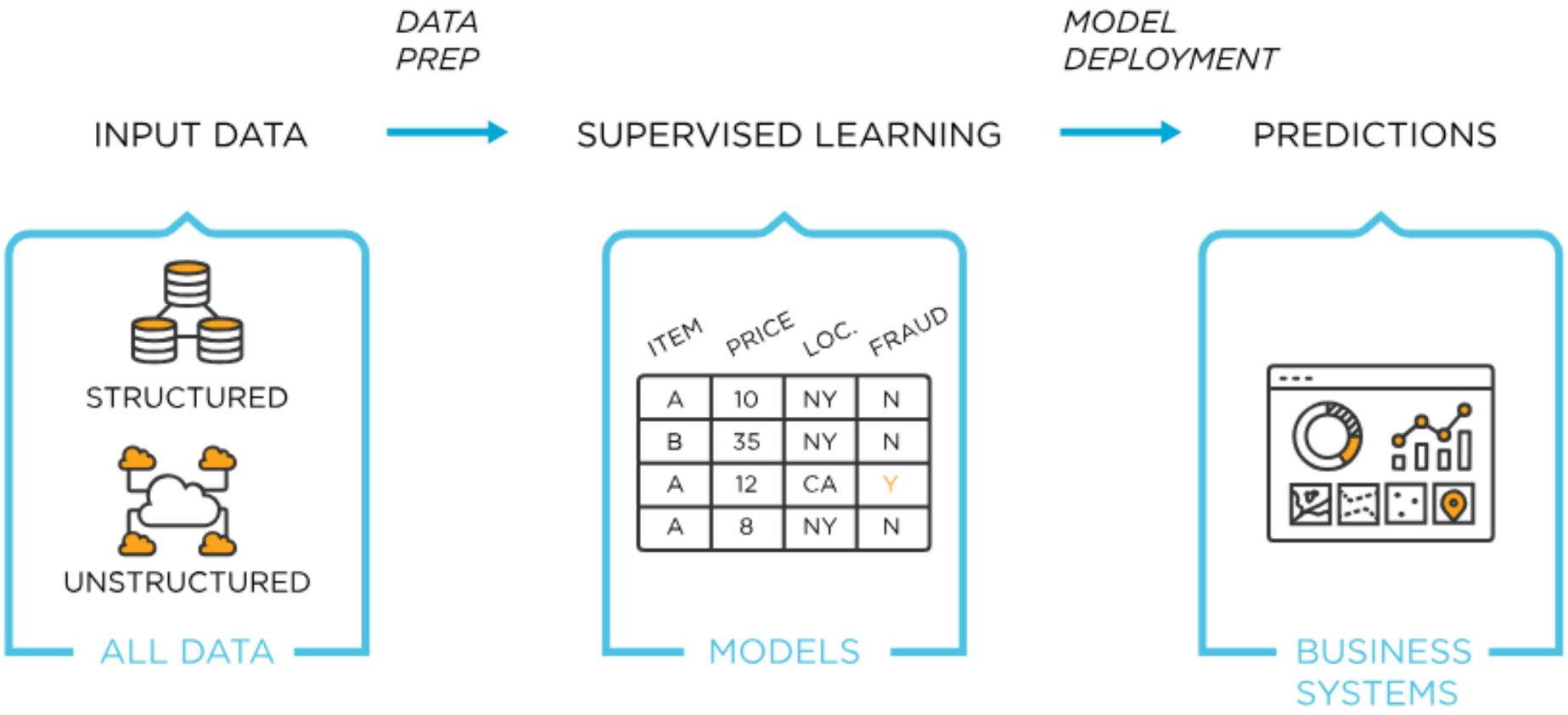
Ans: Data science is the field that combines statistical analysis, machine learning, and programming to extract insights from data.

Q 2. What are the key steps in the data science process?

Ans: The key steps are data collection, data cleaning, data exploration, modeling, evaluation, and deployment.

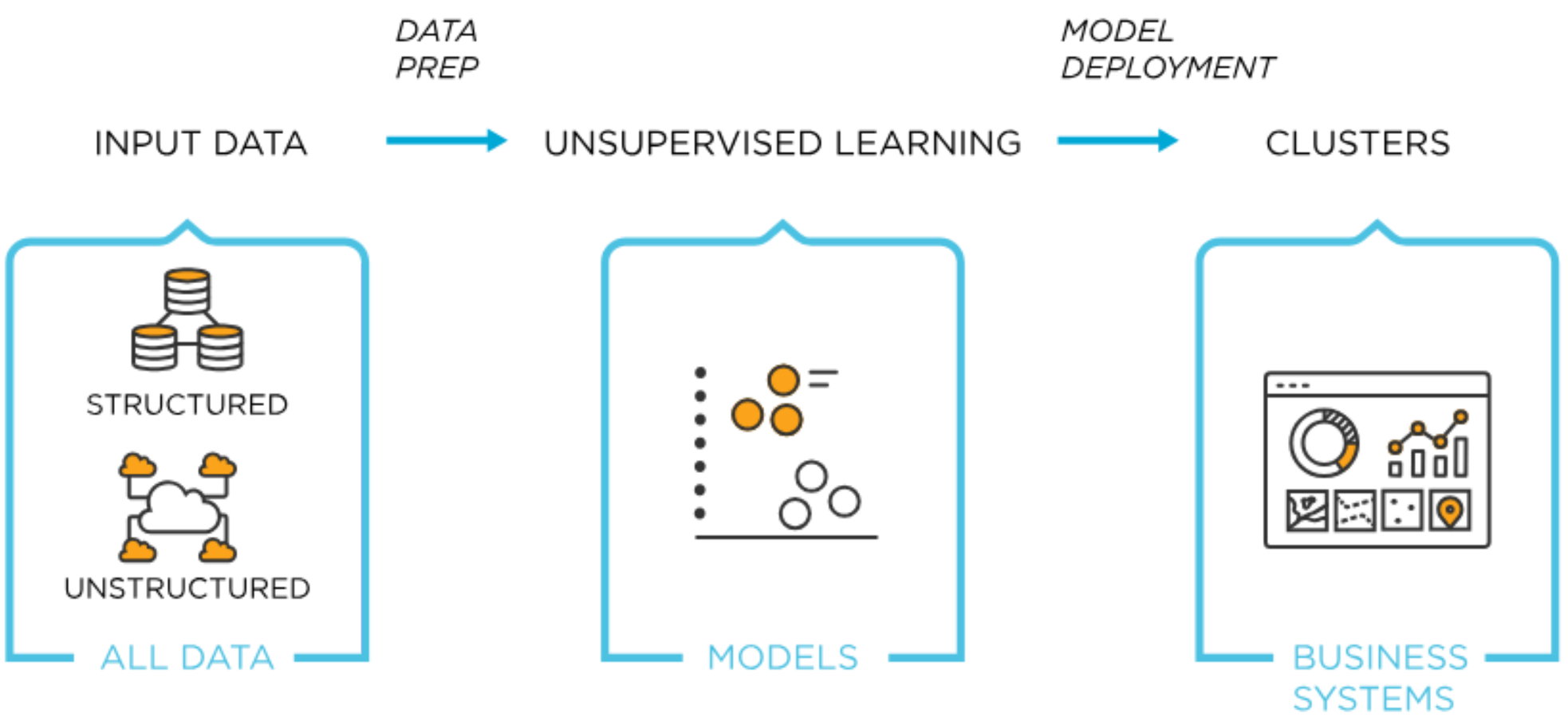
Q 3. What is supervised learning?

Ans: Supervised learning is a machine learning technique where the algorithm learns from labeled data to make predictions or classifications.



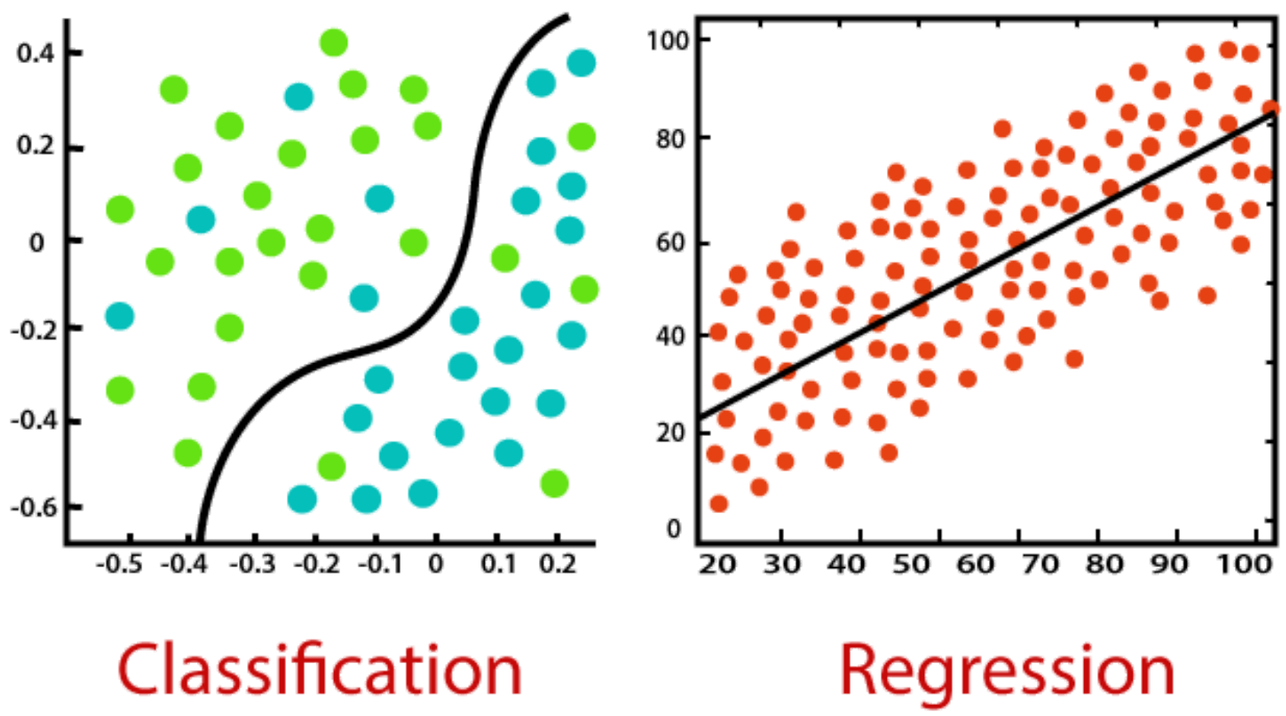
Q 4. What is unsupervised learning?

Ans: Unsupervised learning is a machine learning technique where the algorithm explores patterns in unlabeled data without predefined outputs.



Q 5. What is the difference between classification and regression?

Ans: Classification predicts categorical labels, while regression predicts continuous values.



Q 6. What is overfitting?

Ans: Overfitting occurs when a model performs well on the training data but fails to generalize to new data due to capturing noise or irrelevant patterns.

Q 7. What is feature selection?

Ans: Feature selection is the process of selecting relevant features to improve model performance and reduce complexity.

The three steps of feature selection can be summarized as follows:

- **Data Preprocessing:** Clean and prepare the data for feature selection.
- **Feature Scoring:** Compute scores for each feature to reflect its importance to the target variable.
- **Selection:** Select a subset of the most important features based on their scores, and use them for training the predictive model

Q 8. What is cross-validation?

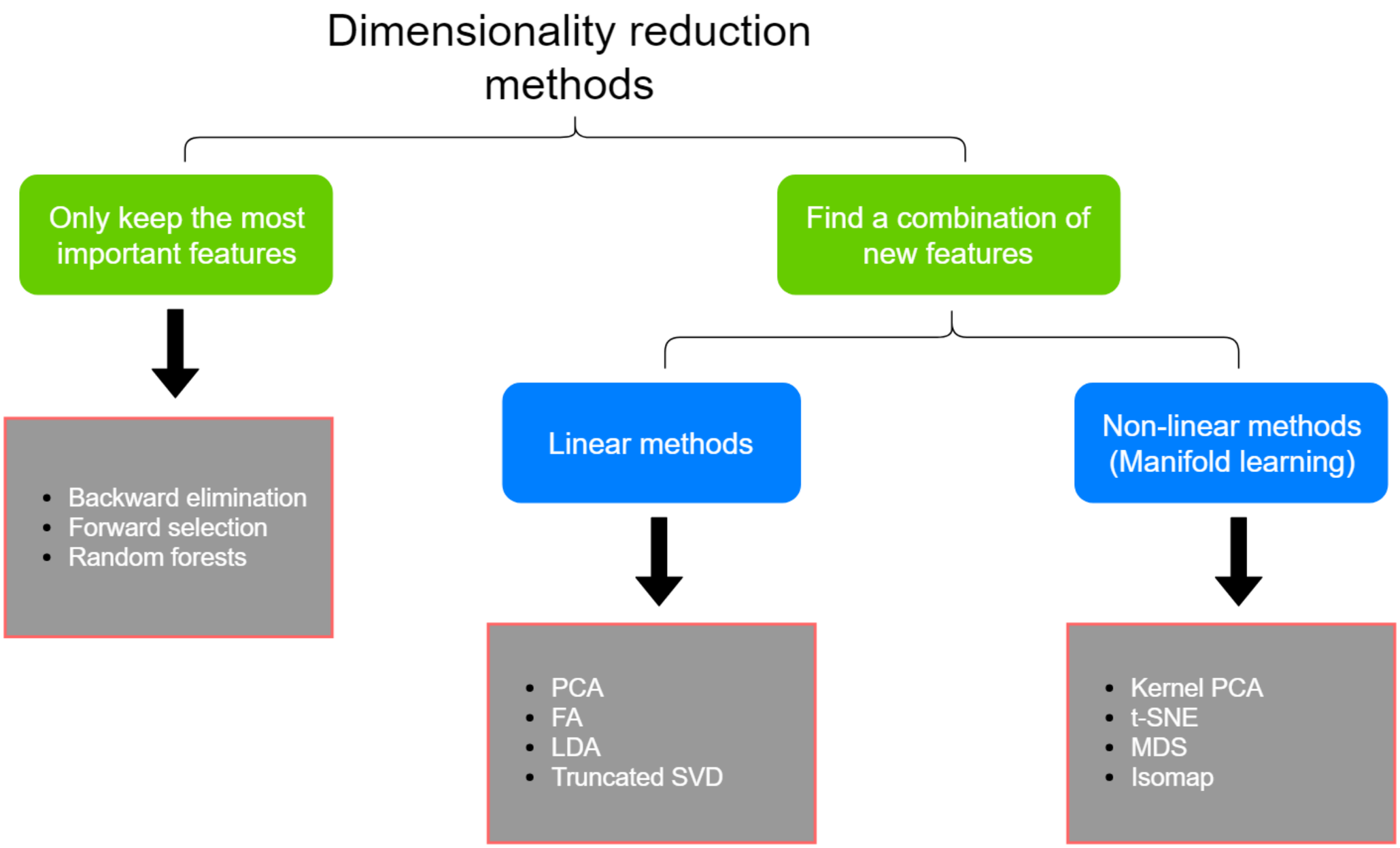
Ans: Cross-validation is a technique to assess model performance by dividing the data into subsets for training and evaluation.

Q 9. What is regularization?

Ans: Regularization is a technique to prevent overfitting by adding a penalty term to the model's objective function.

Q 10. What is dimensionality reduction?

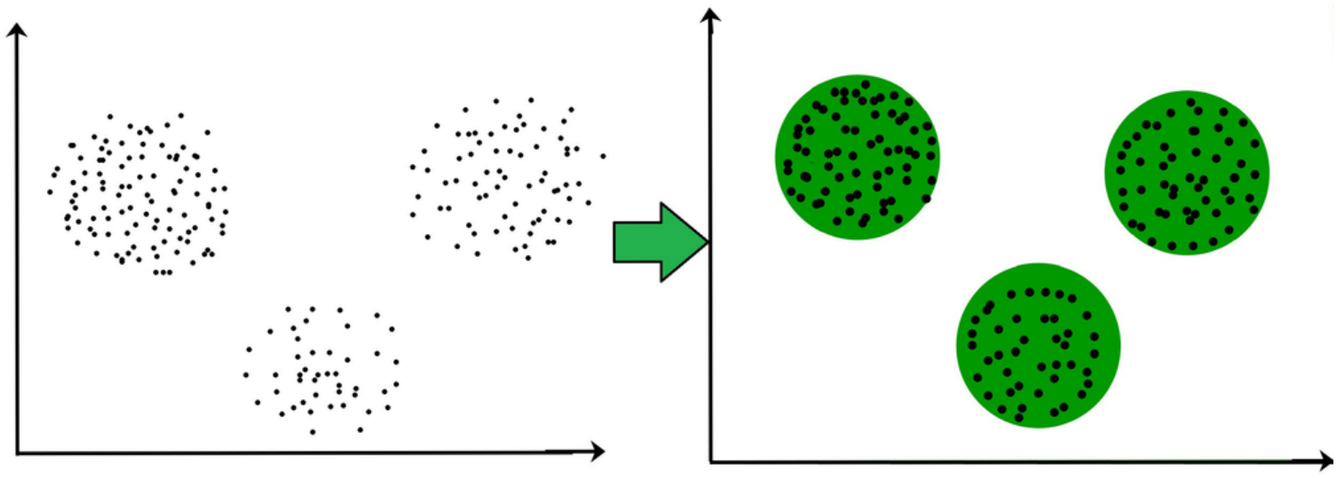
Ans: Dimensionality reduction reduces the number of features in a dataset while preserving important information and structure.



AlgoTutor

Q 11. What is clustering?

Ans: Clustering is the process of grouping similar data points together based on their characteristics or patterns.



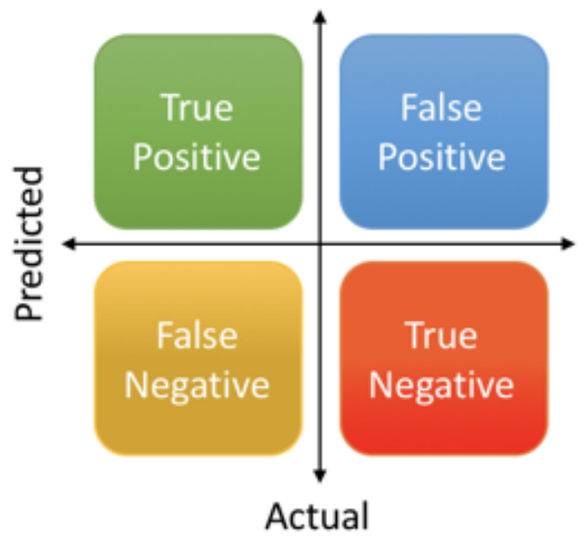
Q 12. What is precision and recall?

Ans: Precision measures the accuracy of positive predictions, while recall measures the coverage of positive instances.

$$\textbf{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\textbf{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

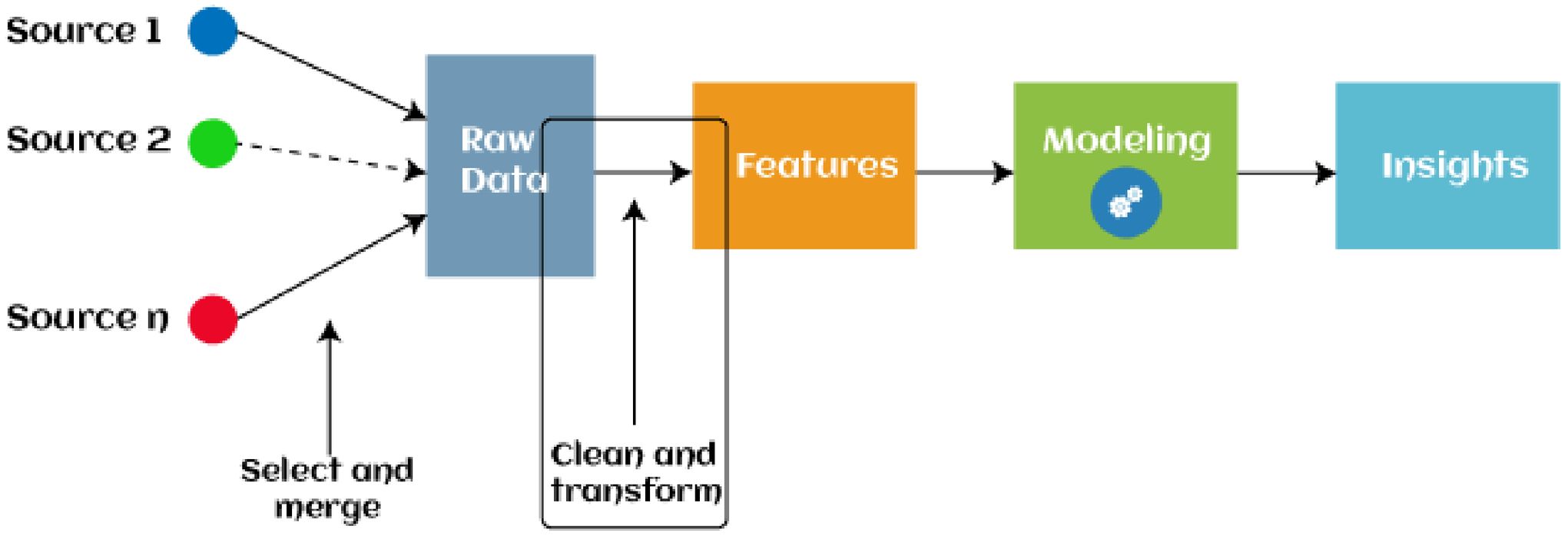
$$\textbf{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$



AlgoTutor

Q 13. What is feature engineering?

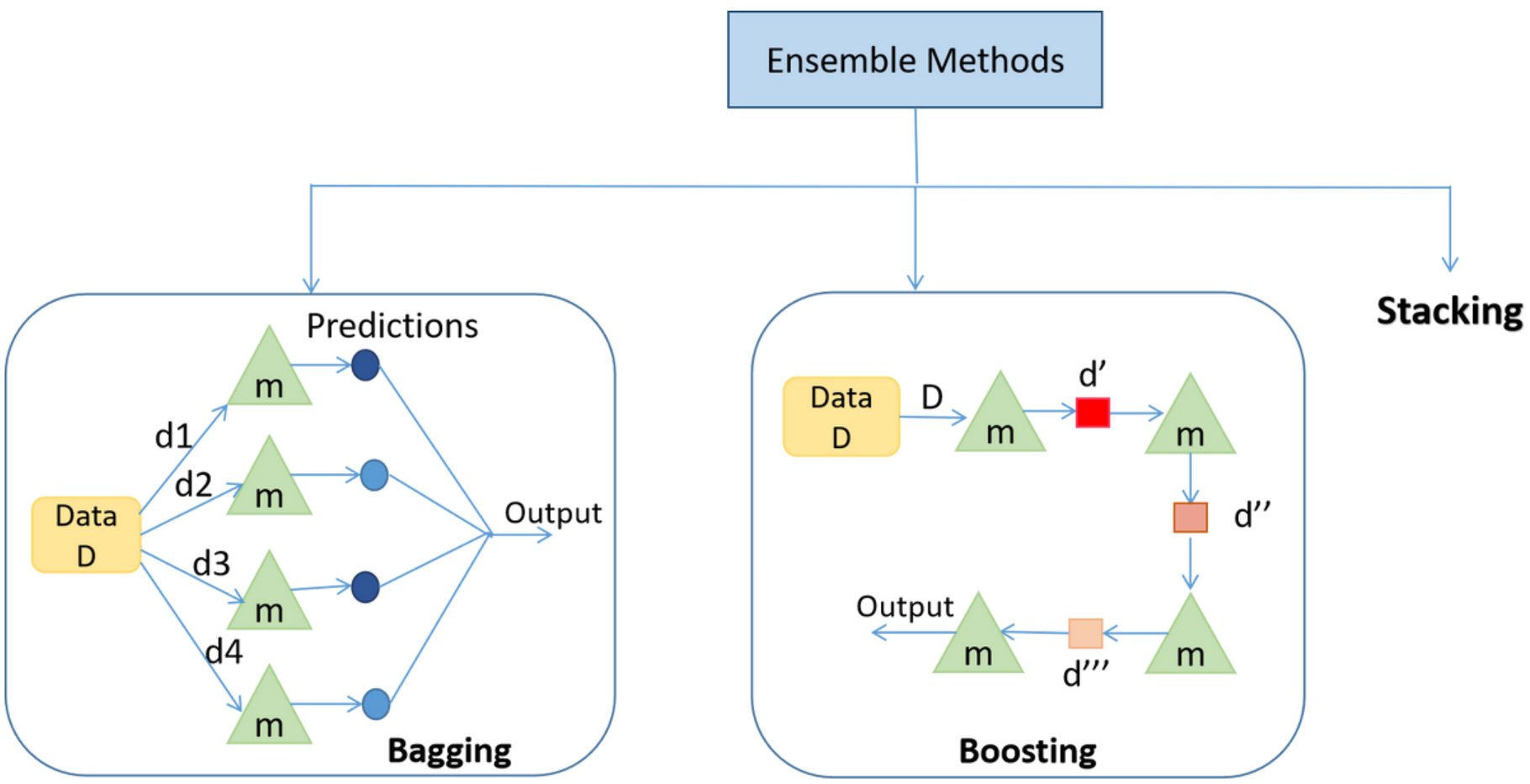
Ans: Feature engineering involves transforming raw data into meaningful features to improve model performance.



Q 14. What is ensemble learning?

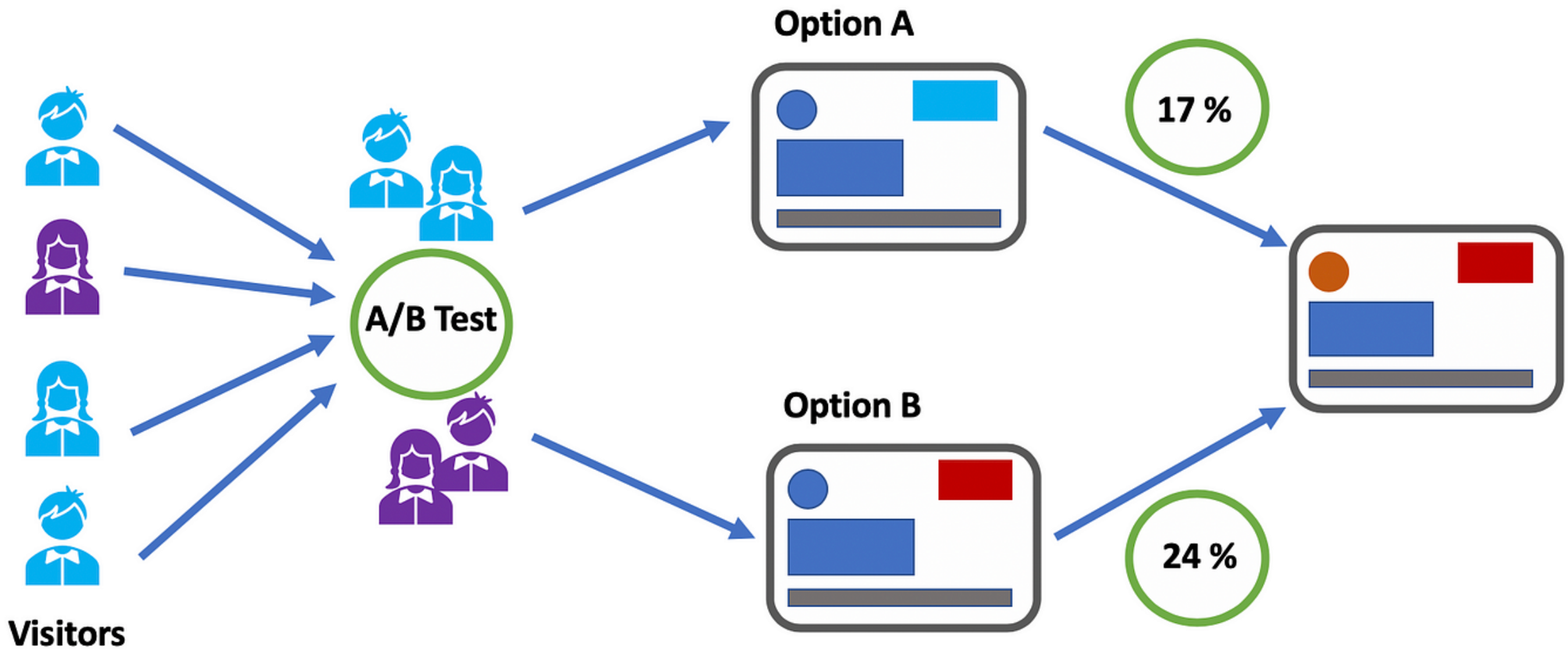
Ans: Ensemble learning is a machine learning technique that combines the predictions of multiple models to make more accurate predictions than any individual model could make on its own.

This is done by training a number of different models on the same data set, and then combining their predictions to get a final prediction.



Q 15. What is A/B testing?

Ans: A/B testing compares two or more versions of a product or feature to determine the best-performing option.



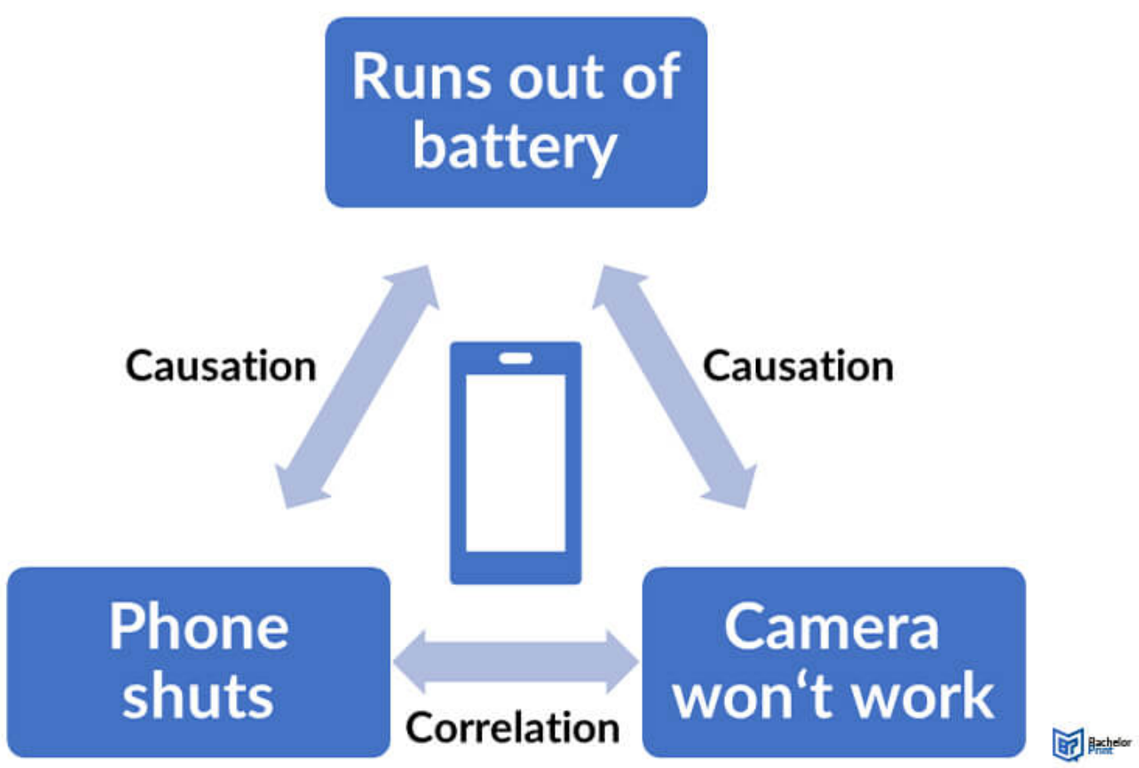
Q 16. What is the central limit theorem?

Ans: The central limit theorem (CLT) is a theorem in probability theory that states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

AlgoTutor

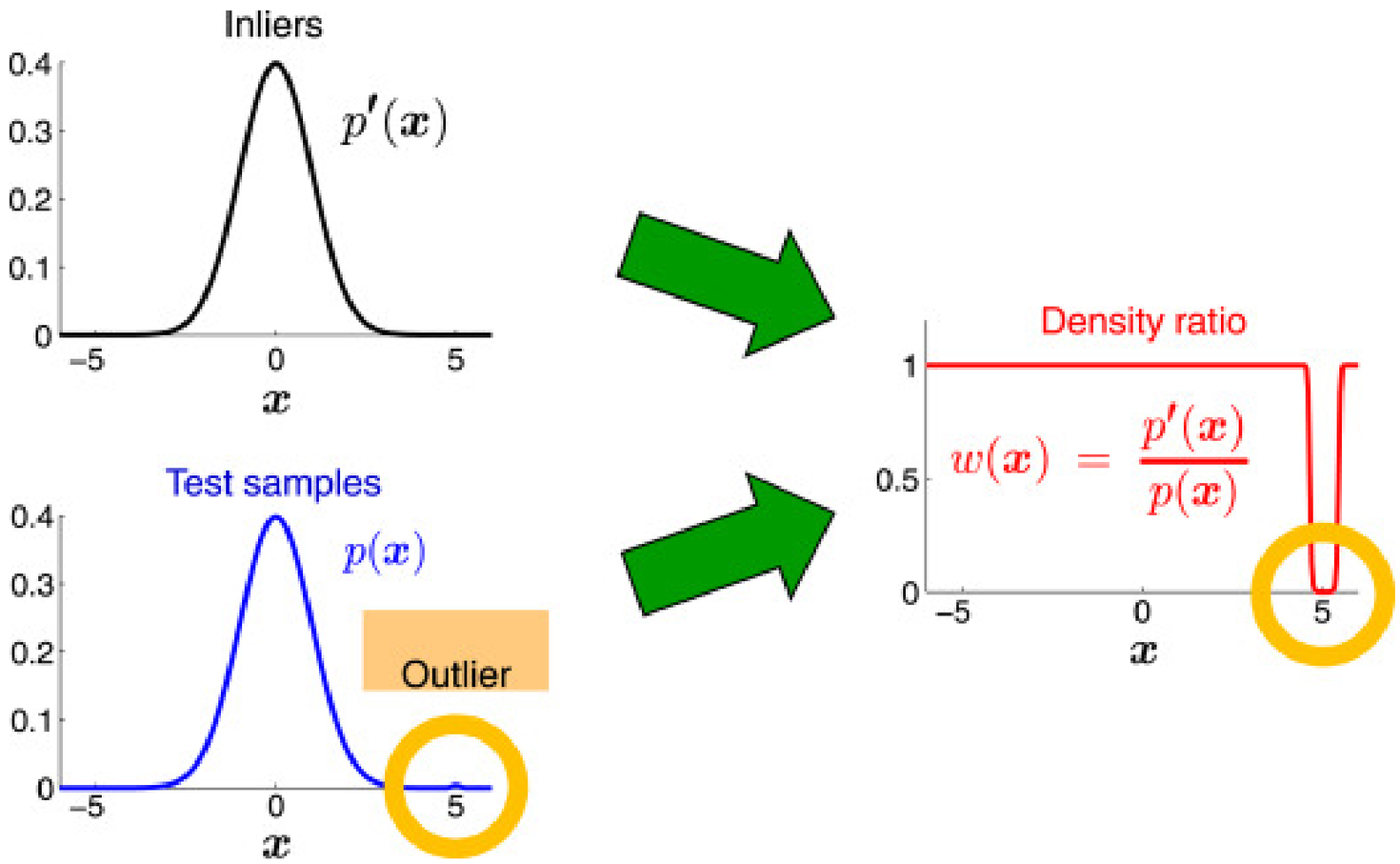
Q 17. What is the difference between correlation and causation?

Ans: Correlation indicates a relationship between variables, while causation implies that one variable directly affects another.



Q 18. What is outlier detection?

Ans: Outlier detection helps identify data points that significantly deviate from the expected patterns.



Q 19. What is the difference between a parametric and non-parametric model?

Ans:

Feature	Parametric models	Non-parametric models
Assumptions	Assume a specific data distribution	Do not make any assumptions about the data distribution
Accuracy	Can make more precise predictions	Less accurate than parametric models
Robustness	Sensitive to changes in the data distribution	More robust to changes in the data distribution
Complexity	Simpler	More complex
Data size	Need more data	Can work with less data
Computational time	Faster to train	Slower to train

Q 20. What is data mining?

Ans: Data mining involves discovering patterns and insights from large datasets.

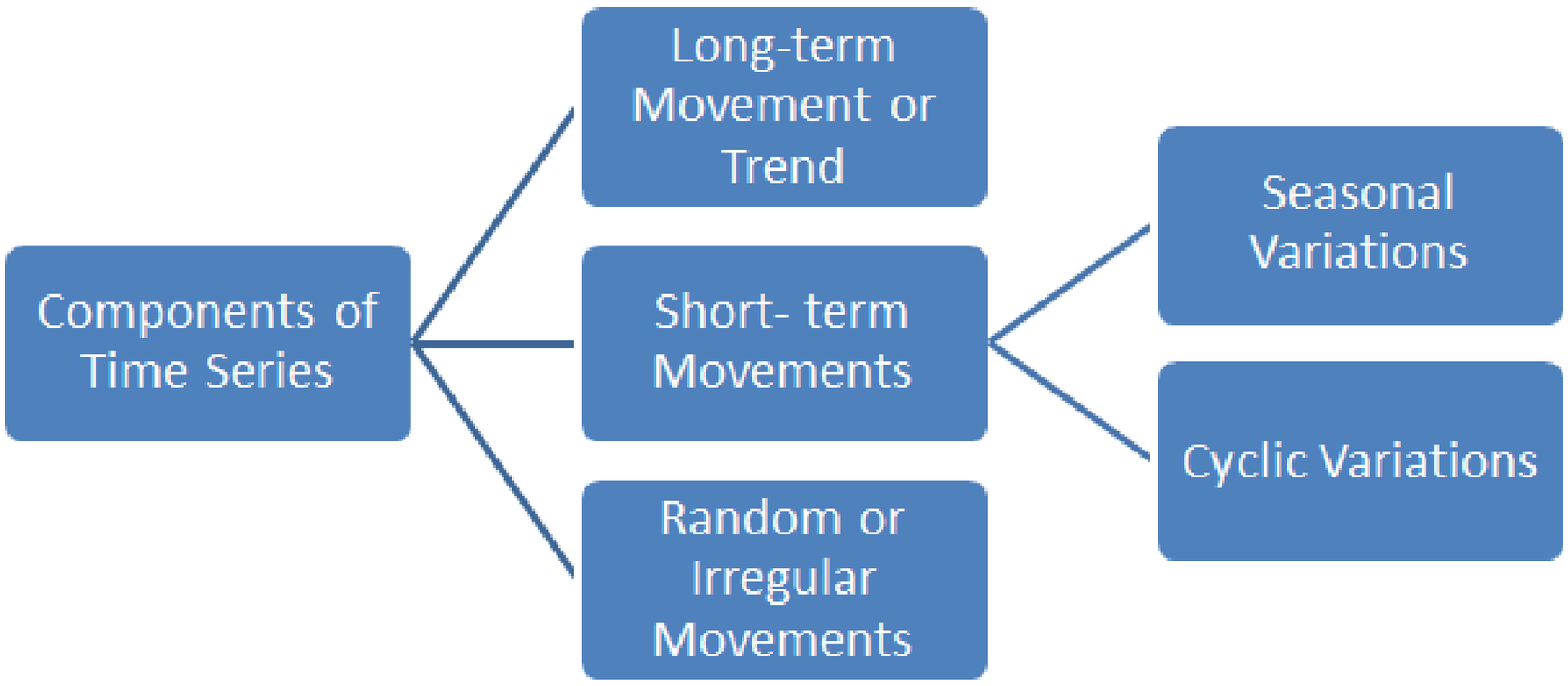
Q 21. What is the bias-variance trade-off?

Ans: The bias-variance trade-off refers to the balance between model complexity and its ability to generalize well to new data.

AlgoTutor

Q 22. What is time series analysis?

Ans: Time series analysis is used to analyze and forecast data points collected over time.

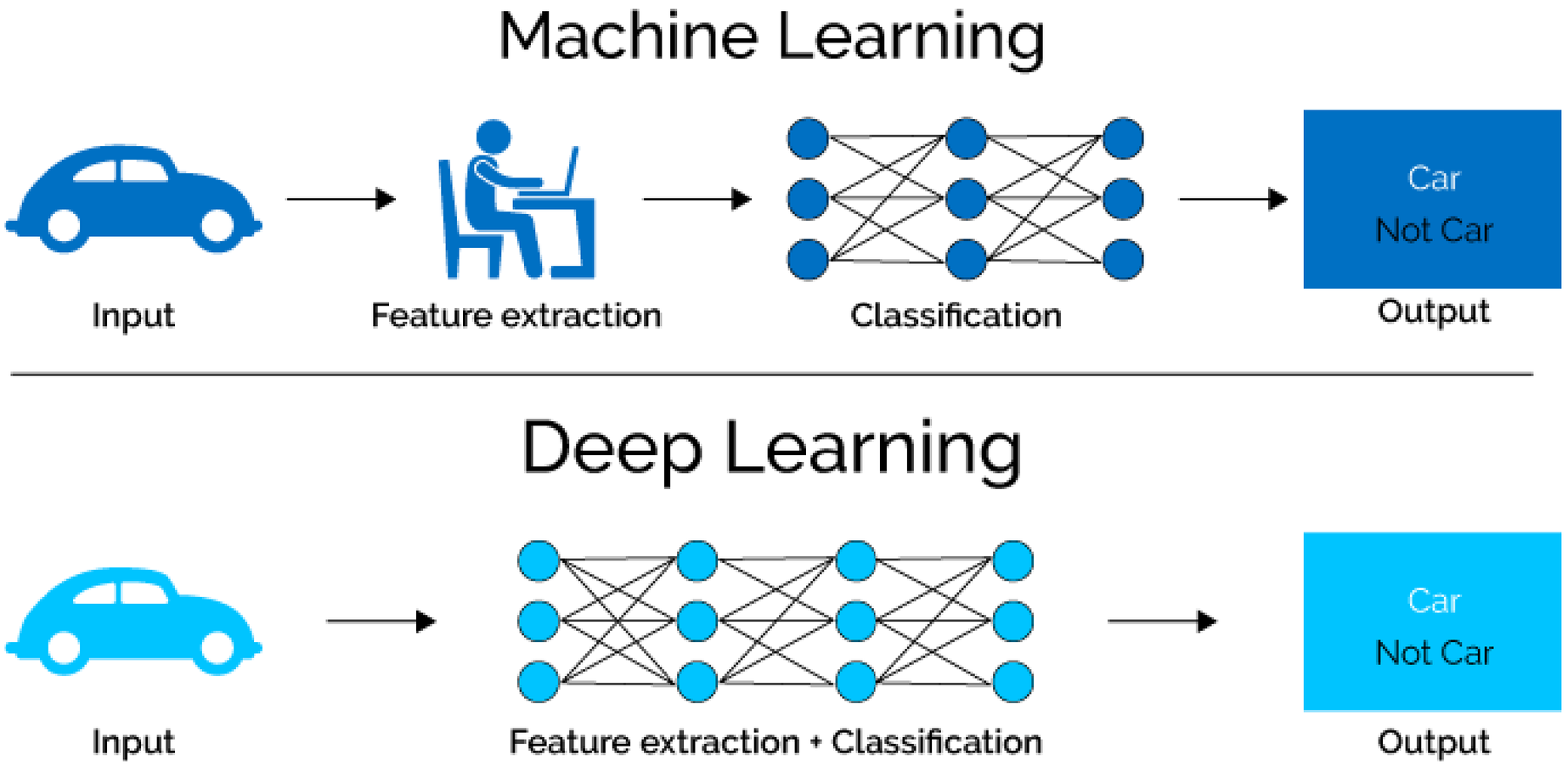


Q 23. What is natural language processing (NLP)?

Ans: NLP focuses on the interaction between computers and human language, enabling machines to understand, interpret, and generate human language.

Q 24. What is deep learning?

Ans: Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers to learn complex representations of data.



Q 25. What is reinforcement learning?

Ans: Reinforcement learning is a type of machine learning where an agent learns to make decisions based on feedback from its environment.

ABOUT US

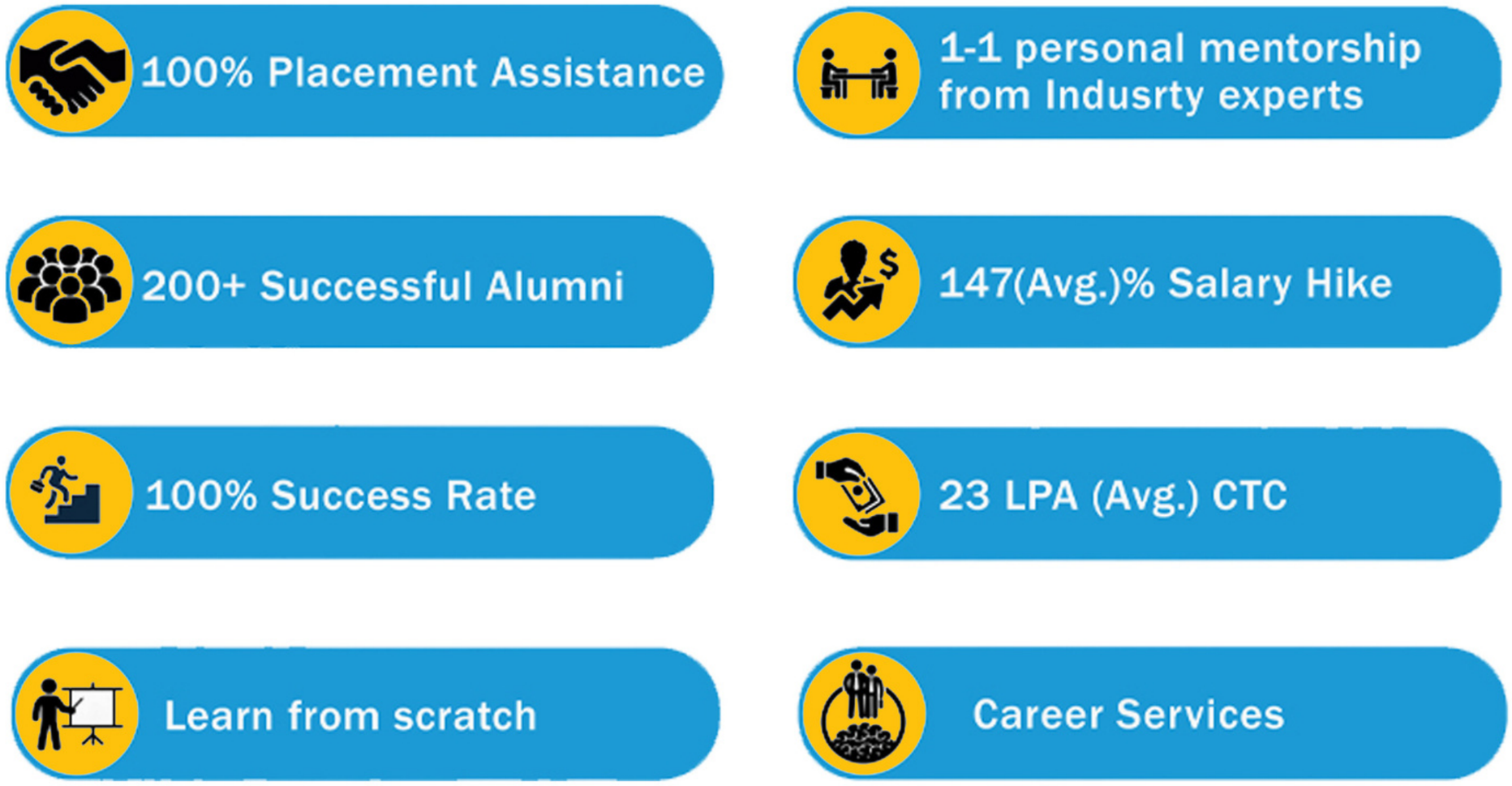
- AlgoTutor is an e-learning platform that provides students with the tools and resources they need to succeed in today's competitive World.

- Our courses are designed by experienced educators and industry experts.

- We offer a variety of learning formats, including video lectures, interactive exercises, and quizzes, so you can learn at your own pace and in your own way.

- AlgoTutor is also committed to providing students with the support they need to succeed. Our team of tutors is available 24/7 to answer questions and provide guidance.

USP of our Programs

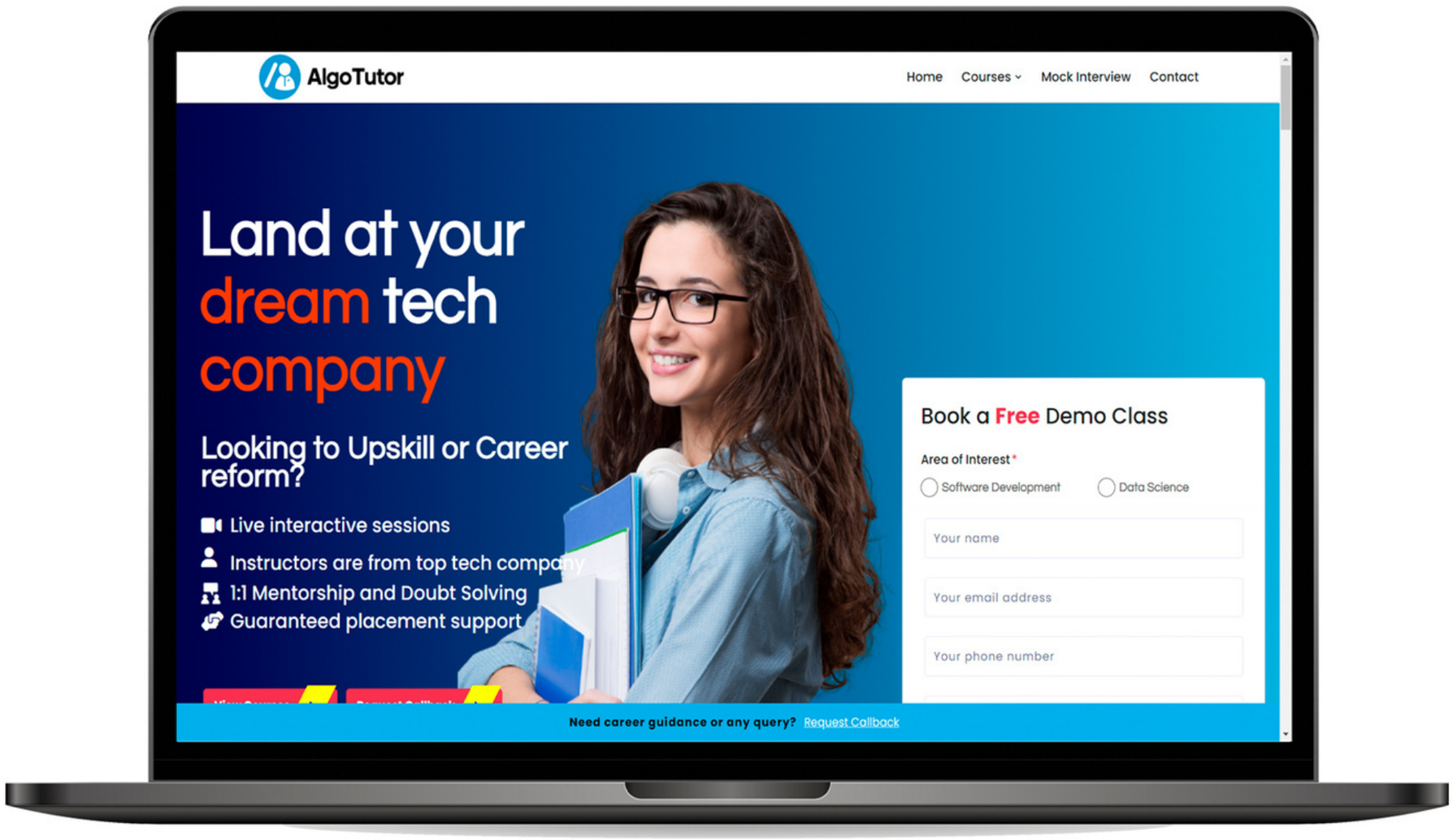


want to Upskill Yourself ?

Explore our Popular Courses

- Data Structure & Algorithms
- Advance System Design (LLD + HLD)
- Advanced Data Science & Machine Learning
- MERN Full Stack Development

!! Visit Our Website !!



+91-7260058093 www.algotutor.io info@algotutor.io